Technical Report

AD 640 124

MECHANIZATION STUDY
OF THE
AEROSPACE MATERIALS
INFORMATION CENTER
(INCLUDING CERAMICS AND
GRAPHITES TECHNICAL INFO CENTER),
WRIGHT-PATTERSON AFB, OHIO

Submitted to

Defense Supply Agency
Defense Documentation Center
Cameron Station, Virginia

by

Booz, Allen Applied Research Inc.
4733 Bethesda Avenue
Bethesda, Maryland 20014

Under Contract No. DSA-7-15489

BAARINC Report No. 914-1-29

September 1966

BOOZ·ALLEN APPLIED RESEARCH INC.

WASHINGTON
CLEVELAND
CHICAGO
LOS ANGELES

TABLE OF CONTENTS

INDEX OF FIGURES

I. SUMMARY

Under contract to the Air Force Materials Laboratory, Wright-
Patterson Air Force Base (W-PAFB), Ohio, the University of Dayton
has indexed and processed some 30,000 documents for the Aerospace
Materials Information Center.   Index terms have been added to computer
tapes for retrieval and printing of the keyword index.

The final choice of an information retrieval system was an in-
verted file system using deep indexing, links, and roles; all material
was processed on the University NCR 304 computer. The initial pur-
pose of the computer operations was the the tabulation of a thesaurus
of the system's vocabulary and the generation of a keyword index that
could be used by individuals in performing manual literature searches.
The magnetic tape from which the thesaurus and dictionary were gen-
erated also provided the opportunity to perform mechanized searches
as the need arose.

Within the past year, all data in the system have been added in
sequential format and without roles to the IBM 7094 at W-PAFB, which
is now used for searching, though the capability still exists at the

-1-

University of Dayton. In addition to searches, the IBM 7094 produces the KWIC index section of the "Materials Information Bulletin."

Currently all documentation received at, or issued by, the Ceramics & Graphites Technical Information Center is included in the overall program developed and operated by the University of Dayton.

All of the 30,000 documents on file at the Aerospace Materials Information Center are unclassified. The collection is growing at the rate of 400 per month. Reference use of the collection and abstracts has averaged about 300 per month, with loans (approximately 600 to 700 annually beside use in the library) being made to Materials Laboratory personnel and in a few instances to people located outside Wright-Patterson AFB.

A weekly "Materials Information Bulletin" announces to potential users all new additions to the collection; the "Abstracts of Active Contracts" announces current projects of the Laboratories. Bibliographies are prepared as requested. The majority of requests are by telephone, though requests by memo or letter are received quite frequently. Fifty percent of all requests are made by DoD contractors; the remaining 50 percent cover requests from various government activities.

II. MECHANIZATION

1. CHRONOLOGY

In early 1960, the Materials Information Branch was established as a centralized technical information service for the Air Force Materials Laboratory. A processing group was organized to assist in the problem of storage, retrieval, and use of information dealing with the materials R&D area. From this original group, the Air Force Materials Information Centers (AFMIC) have evolved. The Aerospace Materials Information Center and the Ceramics and Graphites Technical Information Center are two of the eight information centers. (Appendix A illustrates the eight centers.)

Contract with the University of Dayton was let December 1, 1960, to perform indexing and develop a computer-assisted information system.

Actual work on the contract with the University of Dayton began in September 1961. The system was developed and input began.

In mid-1963, the retrieval program became operational. The program until this time consisted only of input. The thesaurus was revised to retain as much specificity as practical while using a fragmentation system for organic chemicals and applying a modified but similar approach to other areas where an uncontrolled increase in vocabulary appeared detrimental. The established controls have provided a fairly stable vocabulary and eliminated the need for continued revision of the computer tapes.

In 1964, the retrieval program was rewritten for the IBM 7094 at W-PAFB. In this transition, the inverted file was changed to sequential. The data base was added in September although the University of Dayton still formulates the search questions. Thesaurus maintenance, keyword index printing, and data-base updating are still being performed on the NCR 304 at the University of Dayton.

After a year of setting up the program on the IBM 7094, the retrieval program became operational in late 1965.

2. MAJOR PROBLEMS IN DEVELOPMENT

Problems with the free vocabulary began developing within a couple of years, initiating a redevelopment of terms within the system and the establishment of vocabulary controls.

Indexing is usually a chore, although some relief is given when the indexer formulates search questions as well as indexing. Plans call for indexers to supervise students in the indexing and spend part of their time doing more interesting and varied work.

The University's experience has shown negative aspects to the system such as the difficulty encountered in applying role indicators to the terms. After a thorough evaluation, the assigning of role indicators was dropped. It was also discovered that a large amount of time could be saved in the editing operation by letting the computer, rather than the editors, do the cross-relating of terms.

Another problem developed is the indexing of progress reports, which increased the number of overall entries to the system. This procedure was changed; currently, if a progress report is received in the Center and it is the first to be received on the contract, it is assigned a new accession number and processed for indexing. If it is not the first, the document is assigned the same accession number affixed to the first report and filed. Because of this method of handling progress reports, the exact size of the system is indeterminate.

3. THESAURUS DEVELOPMENT

The use of deep indexing with links and roles formulated the basis of this system. The role indicators used were a slightly modified version of those used by the DuPont Company. The first 500 documents that were indexed averaged over 80 link-to-role-term combinations per report. As the information analysts gained experience, the average number of entries per document was reduced to slightly over 40. The elimination of roles from the system reduced the average to about 27 entries per document.

In 1963, when controls were established for the naming of materials, the 18,000 items in the system were reduced to about 10,000. Within the past year, some 6,000 documents being added to the system have generated very few new terms.

Twelve to thirteen terms are assigned for each document, but because of the hierarchial relationships, a document may then be posted to about 20 to 25 terms automatically.

Maintenance of the thesaurus is described in Appendix B.

4. DESCRIPTION OF PROCESSES

Until 1960, a collection of Materials Laboratory's and its contractors' reports had been maintained manually. Reports were logged in, and a report number index card maintained. No subject cross-referencing was available.

Since the beginning of the University of Dayton contract, material added to the computer also has had abstract cards prepared for it. Currently, these abstract cards are maintained at the Materials Laboratory by contractor, contract number, agency, title, and accession number. Card sets are also maintained at the University of Dayton and some of the divisions within the Laboratory.

The computer programs developed and operated by the University of Dayton permit computer searches and also produce a thesaurus and a printed index of the entire collection (see Figure 1 for input flow). Searches and printing of the KWIC index are performed from the same data base using the IBM 7094 at the Materials Laboratory. In order to provide information required to produce the outputs just mentioned, the following procedures are followed:
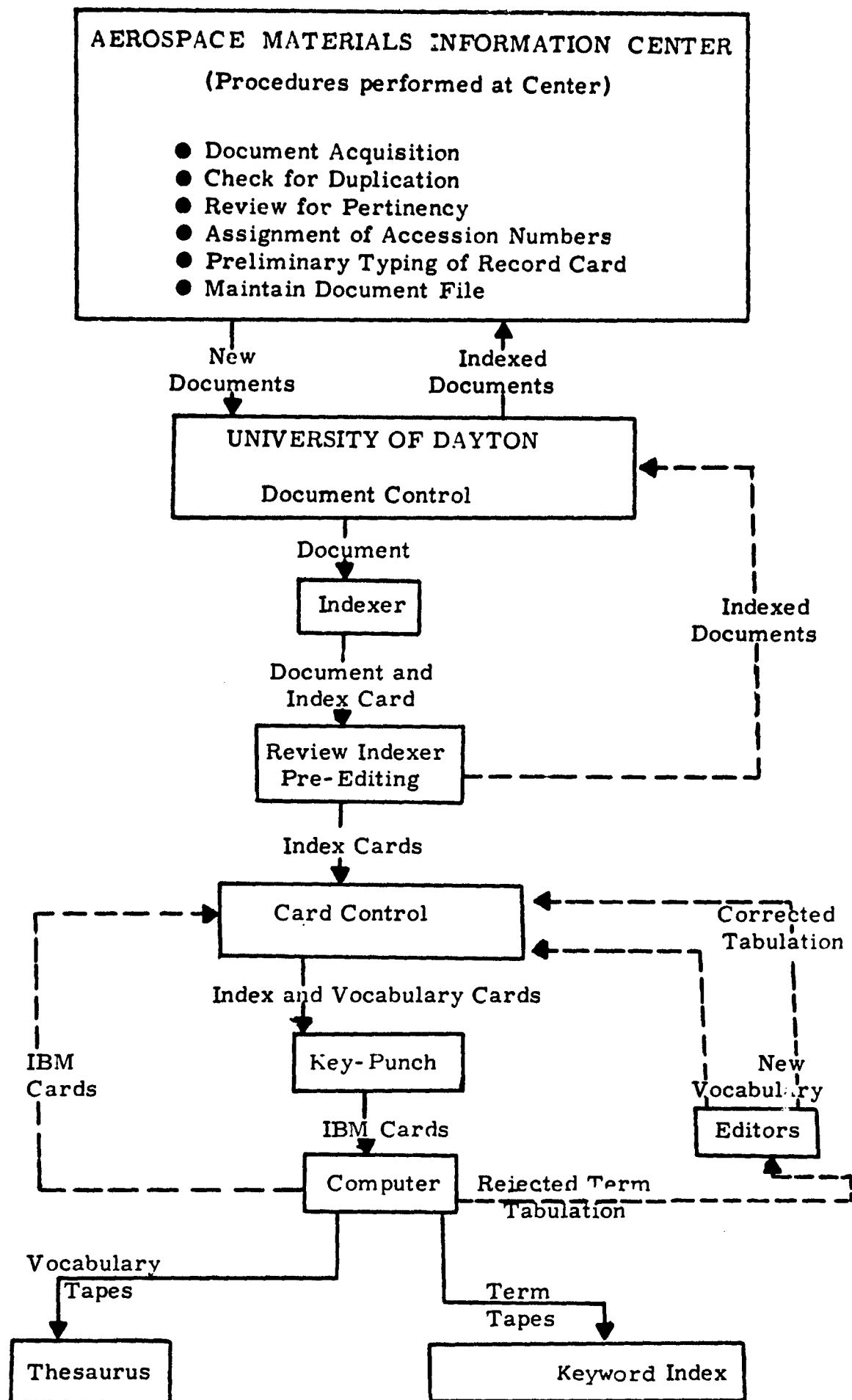
AEROSPACE MATERIALS INFORMATION CENTER

(Procedures performed at Center)

- Document Acquisition
- Check for Duplication
- Review for Pertinency
- Assignment of Accession Numbers
- Preliminary Typing of Record Card
- Maintain Document File

New Documents

Indexed Documents

UNIVERSITY OF DAYTON

Document Control

Document

Indexer

Indexed Documents

Document and Index Card

Review Indexer Pre-Editing

Index Cards

Card Control

Corrected Tabulation

Index and Vocabulary Cards

IBM Cards

Key-Punch

New Vocabulary

Editors

IBM Cards

Computer

Rejected Term Tabulation

Vocabulary Tapes

Term Tapes

Thesaurus

Keyword Index

FIGURE 1

Input Flow of Information

(1)   Input Procedures

1.   Document is received, checked for duplication, and reviewed for pertinency to the collection.

2.   Accession numbers (six digits) are assigned by the University of Dayton staff located at the Materials Laboratory.

3.   Preliminary record card is typed showing accession number, title, date, and contractor.

4.   Documents are forwarded to the University of Dayton for indexing and processing.

5.   Final abstract card is typed including complete bibliographic information.

6.   Indexer indexes the document using an indexing worksheet (referred to as index card; see Figure 2). This index card includes the accession number (access number), title of the report, author or authors, indexer's name and date, and the terms and appropriate links assigned. A link is a unique alphabetic suffix added to the accession number. Terms are selected from the thesaurus by following certain ground rules. See Appendix C for a further explanation of these rules.

Access No. 2 2 F 4 6			Title CHARACTERISTICS OF
PHOTON - COUPLED SYSTEMS			
Author (s) 1. WUNDERMAN			
Indexer RBS		Date 1 MAR 65	
L	L	L	Terms
Z			D- 0465
Z			STANFORD UNIVERSITY
A			THEORETICAL ANALYSIS
A			SEMICONDUCTORS
A			ELECTRONIC COMPONENTS
A			GALLIUM ARSENIDE
A			DIODES
A			PHOTONS
A			EMISSION
A			DETECTORS

FIGURE 2
Index Card

7.   Document and index card go to the review indexer for pre-editing.

8.   Indexed documents are returned to the Materials Laboratory.

9.   Index and vocabulary cards are keypunched.

10.   Cards are processed on the computer; terms not already on tape are rejected.

11.   Rejected terms are sent to editors for review. Terms are held for further review or added, if usage has shown need for the term. Modifications to the various sections of the valid vocabulary are made as described in Appendix B. Figure 3 shows the various card formats.

12.   Final abstract cards are returned to the Materials Laboratory and filed.

13.   Final record cards are filed by the University of Dayton staff, including abstract and index cards.

(2)   Outputs

1.   Thesaurus

The thesaurus contains some 10,000 terms, printed in numeric sequence by numeric code (see Appendix D-1 for sample). Generic Terms (GT), Related Terms (RT),

THESAURUS

D	Numeric	Blank	Alpha
C	Numeric	Blank	Alpha
F	Numeric	Numeric	Alpha
F	Numeric	OOOOOOO	Blank

1-7	9-10	12-18	21-80
Numeric	SA, PO, GT, RT	Numeric	Alpha (May be ERASE)
Numeric	OO (Numeric)	Blank	Alpha (May be ERASE)

GENERIC TAPE

1	4-10	14-20
1	Numeric	Numeric
A	Numeric	Numeric
D		

POSTING INDEX

ADDITIONS

1-6	7-8	11-70
Numeric	Link (Alpha)	Alpha (Punch as briefs)

DELETIONS

1-7	12-19	21-27
Numeric	Numeric & Alpha (Access # & Link)	Blank

ACCESS NUMBER FLIPS

1-7	12-19	21-27
Numeric	Numeric & Alpha (Access # & Link)	Numeric

FIGURE 3
Card Formats

and Posted On (PO) terms are listed immediately below the main term entries with the code GT, RT, or PO. "See" terms and "see also" (SA) terms are included in the thesaurus.

2. Keyword Index

Published annually, with supplements as 4,000 or 5,000 entries are accumulated, this index (Appendix D-2) provides a manual searching tool of all items in the collection. Each term is shown with all the accession numbers posted to it. The accession numbers may then be used to locate the abstract card with full bibliographic data or the document itself. To simplify this manual search, accession numbers are printed in 10 columns corresponding to the terminal digit of the number. The index is kept at the laboratory for quick ready reference-type searches not requiring computer processing.

The index as well as the thesaurus is produced on the NCR 304 at the University of Dayton. Frequency of use can be determined from the index printout but not from the thesaurus.

3. Document Search

Search requests are received by the Materials Labo-
ratory and sent to the University of Dayton for processing.
The search writer formulates the question and writes up the
search in English (Figure 4). The word numbers are usu-
ally looked up by one of the student employees. The codes

IR WORKSHEET FOR SEARCH #_____

PAGE ___ OF ___

?DER	CONNECTOR			WORD	WORD NUMBER
	AND	OR	NOT		

FIGURE 4

are transferred to a keypunch form (Figure 5) used in punch-
ing cards for the retrieval program on the IBM 7094 at the
Materials Laboratory.

SEARCH			
1	2	3	4

SEARCH TITLE
11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40

NAME OF REQUESTOR
41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60

CUTOFF
64 65

MAX #
68 69 70
A.L.L

H
80
H

ORDER	CONNECTOR			WORD NUMBER	ROLE
	AND	OR	NOT		
6 7	8	9	10	11 12 13 14 15 16 17	18 19
01-41					
02-42					
03-43					
04-44					
05-45					
06-46					
07-47					
08-48					
09-49					
10-50					
11-51					
12-52					
13-53					
14-54					
15-55					
16-56					
17-57					
18-58					
19-59					
20-60					

ORDER	CONNECTOR			WORD NUMBER	ROLE
	AND	OR	NOT		
6 7	8	9	10	11 12 13 14 15 16 17	18 19
21-61					
22-62					
23-63					
24-64					
25-65					
26-66					
27-67					
28-68					
29-69					
30-70					
31-71					
32-72					
33-73					
34-74					
35-75					
36-76					
37-77					
38-78					
39-79					
40-80					

FIGURE 5

Resultant printouts (Appendix D-3) indicate the search number, search title, requester's name and extension, special instructions on the search, order of terms, corrections, term numbers, and terms to be searched. Results are in the form of accession numbers for documents containing all or some of the terms specified in the search question. Accession numbers are printed in 10 columns.

Printouts are reviewed by analysts at the Center before the requester receives a copy. The analyst may indicate items which seem to be best suited to the requester's needs, ones which look doubtful, or ones which are ruled out immediately. Local users will receive a printout of terms; outside requesters will receive a list of references or the actual information they required.

4. <u>KWIC Index to "Materials Information Bulletin"</u>

Weekly, the IBM 7094 prints a KWIC Index of new documents added to the collection (See Appendix D-4 for sample). Permuted titles refer to the accession number (found at the right-hand side of the page for each title). This accession number may be used to locate either the abstract or document. A request form (added two months ago) is provided at the back of the bulletin, allowing the user to

circle any accession numbers pertaining to documents for which he wishes to see an abstract. This form is returned to the Materials Laboratory, and abstracts are furnished as soon as possible.

5.   Other

A number of printouts are made to assist in system management. The vocabulary may be printed in alphabetic or numeric sequence. The Generic File may be printed, creating a listing in the same format as the thesaurus, differing only in that all terms at all levels appear. Also, lists of errors are printed by each of the updating runs for correction.

5.   ACTIVITIES BEING PLANNED OR DEVELOPED FOR MECHANIZATION

It is hoped that the "Abstracts of Active Contracts," now being produced manually, can be printed from the DD Form 1498 file at DDC, directly from the computer to reproducible mats.

A Miracode system just purchased will be used to store Materials Laboratory information as well as the "Commerce Business Daily" entries and some AEC materials information being purchased on film from the AEC.

Microfilm will be obtained on all other older documents, micro-fiche on the more current.

III.   PROGRAM SYSTEM DATA

1.   FILES

(1)   Vocabulary File (Master Word List)

This is the basic file of valid descriptor terms and their numeric code equivalents and also contains "see" terms. Two versions of the file are kept, one in numeric sequence which is nearly alphabetic because of the code assignment technique, and another in computer alphabetic sequence which is a slightly different sequence. These and the Thesaurus File are the only tapes in the system that contain English.

(2)   Generic File

This file contains records that reflect the hierarchical relationships of the vocabulary. It consists of pairs of term code numbers, the first being posted on or generically lower than the second.

(3)   Thesaurus File

This is the complete dictionary for the system. The the-saurus is not required by the computer system except as an in-dexing aid and a search writing aid. It contains the codes, first-order generic terms (both higher and lower), "see" terms, "see also" terms, and related terms. Bombardment and irradiation are examples of "see also" terms; each is a good term but there is overlap in meaning.

(4)   Posting Index (Search File)

This inverted index contains term code and accession num-ber pairs in 32-word records. When it is transferred to the 7094 system, it is reinverted to document sequence.

2.   PROGRAMS

(1)   Input

Data cards containing an English descriptor and the docu-ment accession number are read. The English descriptor is compared to the Vocabulary File. If the term is valid, the code number and accession number pair is written on a work tape.

The Generic File is examined, and all higher order terms are selected for automatic posting. These are also written on the work tape. Contents of invalid cards are printed for reassessment.

(2) Post

The work tape resulting from the Input program is sorted by code numbers and posted in proper sequence in the Posting Index.

(3) Search

Batches of four to six questions (maximum of 92 words per question) from one or more users are read and stored. The term codes of the questions are compared to the Posting Index, and their Boolean relationships are evaluated. Evaluation takes place serially from first term to last and on two levels. AND or NOT terms are on the higher level and OR terms are on the lower level. Thus the truth of a single OR term in a string makes true that section of the AND statement in which it appears.

Provision is made for limiting the total number of document retrievals desired and for selecting the number of true phrases to constitute a hit.

Accession numbers of retrieval documents are printed in numeric order within groups showing the number of hit terms.

Two versions of this program have been written, one for the NCR 304 and one for the IBM 7094.

(4)   Print Thesaurus

The preparation of the Print Thesaurus tape involves matching input time with the Master Word List tape. A special subroutine is required to add new terms to these tapes. A cross-match is made to establish that the same terms are included on both tapes. The Print Thesaurus tape is used to produce multiple copies of the thesaurus off-line.

The thesaurus is printed in numeric sequence. All first-order generic terms, related terms, "see" terms, and "see also" terms are printed.

(5)   Print Keyword Index

English terms are extracted from the Vocabulary File and merged with data from the Posting Index onto a work tape. The tape is sorted into sequence by the English term and accession numbers, and printed off-line. A supplement to the Posting Index

-22-

is printed when 4,000 or 5,000 entries are accumulated. Once each year, all the supplements are merged and a complete version is printed.

(6)   Print Vocabulary-Numeric
Print Vocabulary-Alphabetic

The Vocabulary file of codes and English is printed in the sequence indicated. To date, the alphabetic sequence has not been printed.

(7)   Print Generic

The Generic File is printed in numeric sequence.

(8)   Vocabulary Update

This program makes additions and deletions to the Vocabulary File. An addition is a new term and an unused code number. Code numbers are assigned so that a nearly alphabetic sequence is maintained. Unnecessary terms are deleted from the file. Those terms are also deleted from the Posting Index.

(9)   Flip

This program performs interchange of one term for another. It deletes the first term from the Vocabulary File and inserts the

second. It then moves the postings from the first to this second

term in the Posting Index. If the second term number is zero,

all entries in the Posting Index are simply deleted.

(10) Change

The English portion of an entry in the Vocabulary File is

replaced by the contents of the change card.

(11) Access Number Flip

A single accession number may be removed from posting

under one term and inserted under another term in the Posting

Index to correct mispostings.

(12) Delete Generic

This program deletes single entries from the Generic File.

(13) Add Generic

This program adds single entries to the Generic File.

(14) Thesaurus Update

This program is used to make changes to the Thesaurus

File. Change cards containing code numbers and designations

(e.g., GT, PO, SA, RT, and "see" terms) are read, sorted, checked for validity against the Vocabulary File, and posted. Invalid terms are printed for correction.

(15) Punch

The contents of the Posting Index are punched into cards for transfer to the IBM 7094 for retrieval. This step is necessary because of magnetic tape incompatibility between equipments.

(16) Compare

This program compares the vocabulary to the thesaurus to find terms that exist in the vocabulary but do not appear in the thesaurus. These terms and invalid terms from the Thesaurus Update are reconciled by deleting or by adding in the proper place, to assure one-to-one correspondence in the two files.

(17) KWIC

This is a standard program provided by IBM to all users. It prints a permuted title listing for all new documents.

IV. EQUIPMENT, COSTS, AND EVALUATIONS

1. EQUIPMENT

NCR 304   The NCR 304 system is owned by the University
                 of Dayton and is used primarily for nonscientific
                 data processing problems. In addition to the
                 following equipment, the computer also has
                 paper tape in and out and typewriter control at
                 the console.

    304-B   with 4K memory

    330   magnetic tape controller

8   332   magnetic tape handlers

    320   multipurpose converter

    340   printer

    380   card reader

IBM 514   card punch

IBM 7094-7044   The 7094-7044 system has been rented by the
                 Digital Computations Division, a service facility
                 for all of W-PAFB, for about one year. It is
                 used primarily for scientific computations, and
                 about 85 percent of the programs are in FORTRAN
                 language.

    7094   direct couple system with 32K memory

8   729   Mod IV tapes

                 disk (10 million words)

	7044	input-output computer
	1402	card reader/punch
6	1403	line printers

2. TIME AND COSTS

The University of Dayton personnel assigned to the contract consist of five full-time employees (professional, technical for subject indexing) at the University and five full-time employees located at the Center.

The contract with the University of Dayton has cost the Center approximately $150,000 per year for the last three years. This includes system development, indexing, retrieval statement composition, and file maintenance.

The search program, which ran approximately 30 minutes on the 304, takes only 3 minutes to run on the 7094. Some 400 searches were made during the past year, and some 600 are expected this year. In addition, 10 to 15 manual searches are made each month.

Costs on the 7094 (at $400 per nour) have been approximately $15,400 in the 15 months since September 1964, including $2,300 in November 1965. This includes the cost of printing KWIC title listings

and operation of the system. Since the Center is not charged for search time on the Materials Laboratory's IBM 7094, they feel they save $11,000 by using this computer.

3. **EVALUATIONS**

The information retrieval system established by the University of Dayton became operable in mid-1963. At that time, a thesaurus of approximately 10,000 terms and a dictionary containing references to approximately 6,500 documents were tabluated on the NCR 304 computer, and multiple copies of each were forwarded to the Laboratory. It was only after this date that any form of an evaluation could be made on the propriety of the selected system and its various components.

The use of links and roles had been added to the indexing process in order to reduce the number of false drops and to permit highly sophisticated searches. It became desirable to determine the depth of indexing being performed on each document and whether links and roles were serving the purpose for which they were intended.

(1) <u>Evaluation of Links and Roles in Searches</u>

A study of the value of links and roles was undertaken by a student of the Air Force Institute of Technology (AFIT) as his graduate thesis. Working with the University of Dayton staff, he

ran actual searches on which to base his evaluation. In this study, roles were found inferior to links. The use of links showed a reduction of irrelevant information by over 56 percent while incurring less than five percent loss of relevant information. The study recommended the dropping of roles, or an increase of the word limit for computer searches from 40 to 100 (40-word limit is sufficient when using links alone). The recommendation to drop roles was accepted, recognizing that it was almost impossible to provide definitions which would assure the use of the same role in a question as in the original indexing. Complete results of this study are given in ML TDR 64-152 (see Reference 1).

A second evaluation of links and roles was made by a University of Dayton employee experienced in setting up actual search questions. This study supplemented the first and confirmed the decision to drop roles.

(2)   Special Indexing Studies

A special study has been made of the indexing process. This study was devoted to the indexing of documents in which attention was paid to the location of the terms indexed. Each document was indexed in four parts: the abstract, summary and

conclusions, table of contents and figures, and the body of the report. The indexing was done in the indicated sequence, and a color code was used to distinguish the indexing of the four parts. The indexing of each subsequent section was used to complement and/or supplement the indexing of the preceding sections. The results of this study are tabulated in Tables 1 and 2 of RTD-TDR-63-4263 (see Reference 2).

Realizing that different disciplines presented distinct indexing problems, the documents were categorized as being in the area of ceramics, chemistry, metallurgy, or physics. However, since the reports did not necessarily contain all four sections from which terms were indexed, i.e., all reports did not have abstracts, etc., additional tabulations were made of those reports having similar sections. In addition, the results of this study show some of the costs added to the system by the use of links and roles; e.g., 50 percent more entries caused by use of links and roles undoubtedly results in increased indexer's time, keypunch expenses for time and materials, additional computer processing time, and additional postings.

(3) Additional Studies

The Center has, for the past three years, run searches on projects proposed for contract and distributed these to the project offices involved. Following these searches, a questionnaire (see Figure 6) was sent to each recipient to determine the value of the particular search. At the end of 1964, some 42 replies had been received; 22 indicated that the search results had been of help, and six indicated that the search results had actually cancelled parts of projected projects. Some 183 searches were performed this past year, but results of the questionnaire are not yet available.

EVALUATION OF INFORMATION RECEIVED FROM

_____

Estimated
Percent

. Material was closely related to my work   ____%

Material was moderately related to my work   ____%

Material was remotely related to my work   ____%

Material was totally unrelated to my work   ____%

. Material was new to me   ____%

Material was not new to me   ____%

Receiving this material affected the course
   of the work   [____]

Receiving this material affected the choice
   of contractor   [____]

This material indicated some anticipated work
   was unnecessary   [____]

This material made no contribution to my work   [____]

would have appreciated
Journal Literature   [____]   Patent Literature [____]   _____ [____]

s material adequately pin-pointed?   YES [____]   NO [____]

mments.

FIGURE 6
Search Questionnaire

-32-

R E F E R E N C E S

1. "An Evaluation of Links and Roles Used in Information Retrieval,"
   Technical Documentary Report No. ML TDE 64-152,
   July 1964, AF Materials Laboratory, RTD, AFSC, Wright-
   Patterson Air Force Base, Ohio (AD 606 192).

2. "Establishment of a Coordinate Indexing Retrieval System for the
   Air Force Materials Laboratory," Technical Documentary
   Report No. RTD-TDR-63-4263, March 1964, AF Materials
   Laboratory, RTD, AFSC, Wright-Patterson Air Force
   Base, Ohio (AD 428 423).

APPENDIX A

AIR FORCE MATERIALS INFORMATION CENTERS

A. F. Materials Laboratory

Ceramics & Graphites Tech Info Center

Machinability Data Center

Elect. Prop Center

Aerospace Materials Info Center

Thermophysical Prop Center

Radiation Effects Info Center

Defense Metals Info Center

Mech Prop Center
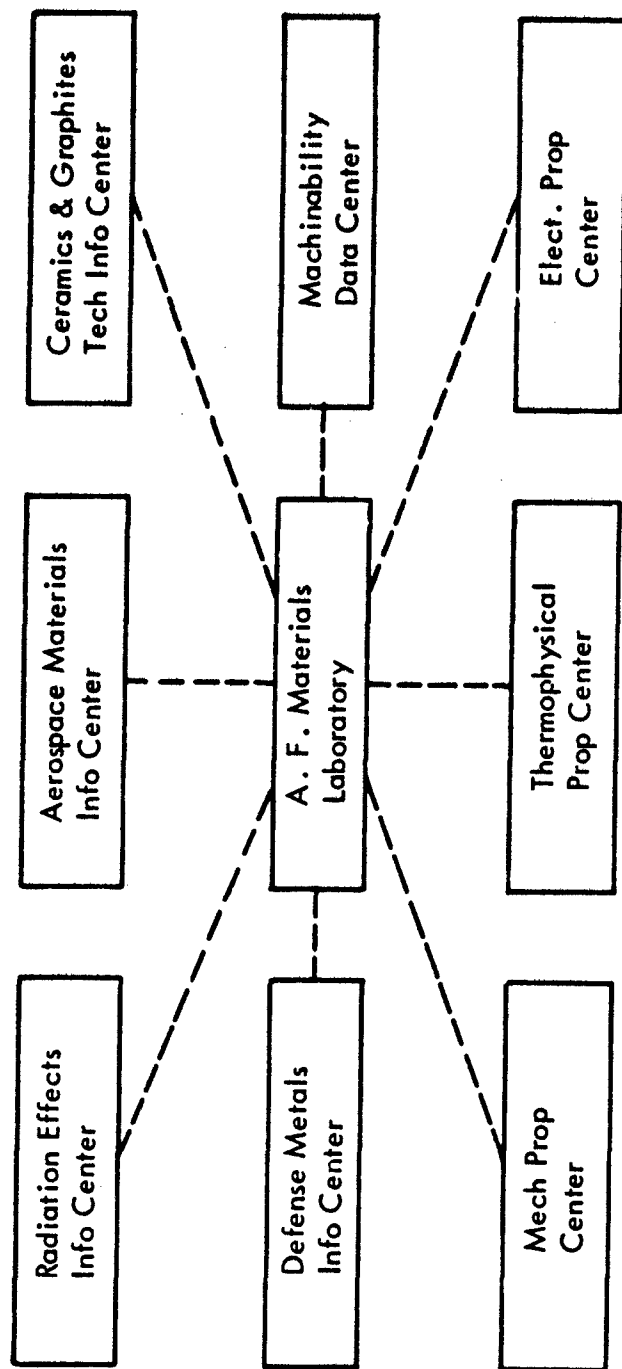
APPENDIX B

MAINTENANCE OF THE THESAURUS

THINGS YOU CAN DO TO THE VOCABULARY

	CODE	COMMENTS
1.	A	Add a NEW word to the system; should have a number which has never been used
2.	C	Change the spelling of an existing number. Effect is to delete one word and add another with the same word number
3.	D	Delete existing word on vocabulary which DOES NOT have postings on the posting index
4.	F	a.) Flip to "zeros" for a delete only--those words which DO HAVE postings on the posting index. Effect is to delete the word from the vocabulary AND the postings from the posting index; new word is NOT added to vocabulary. b.) Flip to another word number for delete and add. Effect is to delete one word from vocabulary and add a NEW word with a NEW number, while "flipping" documents from the old word number to the new word number, deleting the documents from the old word number.

THINGS YOU CAN DO TO THE THESAURUS

COMMENTS

1. Add a new master word # (not previously in the thesaurus) using 00 as the posting.

2. Add SA, PO, GT, or RT to an already existing master word # in the thesaurus. If no master item exists (a 00 item) then program prints error and new card is not added into thesaurus.

3. Delete master word # and ALL corresponding postings using 00 and ERASE

4. Delete individual postings SA, PO, GT, or RT using ERASE. This is used when master word item is to remain in thesaurus with or without other postings.

THINGS YOU CAN DO TO THE GENERIC TAPE

	CODE	COMMENTS
1.	A	Add a NEW PO to the generic tape; both the master word # and the PO word # MUST be in the vocabulary
2.	D	Delete an existing item from the generic tape; there must be a one-to-one correspondence between cards and items you wish to delete; i.e., one card must be punched for each item you wish to delete

THINGS YOU CAN DO TO THE POSTING INDEX

1. Add new documents--word items to the Posting Index (from index cards or error sheets). Cards will be matched with vocabulary (sorted by word) and output will be matched with generic tape. You will get a new error sheet, also, the generic tape match output will be merged with the search tape and a supplement will be printed (or the entire search tape, if desired).

2. FLIPS--the flip cards from the vocabulary update will be used to delete access #'s and add them under new master word #'s. This output WILL be matched with the generic tape, which output in turn will be merged with the search tape. A printout can be made available if you so desire.

3. "Bad" access numbers can be deleted from the posting index. One card per access # must be punched. The "bad" documents will be deleted, leaving good documents untouched--even under the same master word #, i.e., you can delete any number of "bad" documents under a certain word #, leaving "good" documents under that word #. Just remember that one card must be

punched for each "bad" document. Also, some of the deleted documents may belong elsewhere. You can add these documents to the proper word simply by utilizing this feature on the input cards. If additions exist, the output will be matched with the generic tape and in turn will be merged with the search tape. The effect is similar to that of a flip card--the difference being the ability to "flip" certain but not all the documents under a given word #.

APPENDIX C

INDEXING GROUND RULES

APPENDIX C

INDEXING GROUND RULES

The elimination of role indicators simplified the indexing pro-
cedures and reduced the number of indexing ground rules. Specific
indexing techniques are primarily concerned with terminology or vocab-
ulary control while the basic indexing procedures remain unchanged.
First, the information analyst must become familiar with the contents
of a document before he can determine its important concepts. For the
majority of documents, this can be accomplished by reading the ab-
stract and conclusions, and by scanning the table of contents and the
body of the report. The key terms that best describe the report's
contents are then chosen, and links are used to maintain separation of
the terms of principal concepts. The use of links has become valuable
in separating organic compounds and preventing the "connection" of
substituent groups to the wrong basic structure.

Any indexed terms that are not already included in the system's
vocabulary are recorded on master word cards. These cards are re-
viewed once a week by all staff members. This method permits the
revising of the vocabulary tape prior to the updating of the Posting
Index, and at that time, only keypunching errors should be rejected by
the computer.

The changes made in the system's vocabulary made it advisable to make one other change in the indexing procedure. Realizing the possibility that users might ask for information on specific materials, a decision was made to use more scope notes on the index cards in order to simplify the screening of search results. Any notes that an information analyst thinks would be beneficial in the screening process are written on the index card in script and enclosed in parentheses. These notes are not keypunched and, therefore, do not affect the computer programs.

APPENDIX D

SAMPLES OF OUTPUTS

```
3460500    RHEOLOGICAL PROPERTIES
1498000      PC  FLOW PROPERTIES
3148500      PC  PHYSICAL PROPERTIES

3461500    RHODIUM
24 0500      PC  METALS
3430500      PC  REFRACTORY METALS
4376410      PC  TRANSITION METALS

3462000    RHODIUM ALLOYS
3428500      PC  REFRACTORY ALLOYS

3462100    RHODIUM ANTIMONIDE
0349500      PC  ANTIMONIDES

3462200    RHODIUM BORIDE
0424500      PC  BORIDES

3463000    RHODIUM COATINGS
0673500      PC  COATING
2400500      PC  METALS
3420500      PC  REFRACTORY METAL COATINGS
3430500      PC  REFRACTORY METALS
3461500      PC  RHODIUM
4376410      PC  TRANSITION METALS

3463450    RHODIUM SILICIDE
3598000      PC  SILICIDES

3466500    RHODIUM-IR
3428500      PC  REFRACTORY ALLOYS
3462000      PC  RHODIUM ALLOYS

3466500    RHODIUM-OS
3428500      PC  REFRACTORY ALLOYS
3462000      PC  RHODIUM ALLOYS

3467000    RHODIUM-PD
3428500      PC  REFRACTORY ALLOYS
3462000      PC  RHODIUM ALLOYS

3468000    RHODIUM-PT
3428500      PC  REFRACTORY ALLOYS
3462000      PC  RHODIUM ALLOYS

3468500    RHODIUM-RU
3428500      PC  REFRACTORY ALLOYS
3462000      PC  RHODIUM ALLOYS

3470000    RHODONITE
0758000      PC  CALCIUM SILICATE
2330000      PC  MANGANESE SILICATE
2665550      PC  MINERALS
2572500      PC  MIXED OXIDES
3596500      PC  SILICATES

3474500    RIVETS
1440100      PC  FASTENERS

3474500    ROCKET MOTOR CASES
```

PROPYL 23

15000A	10981A	10933A	605H	577C	12619I
	12301A	13733L	125550		

PROPYL 34

		323A			238A
		323B			

PROTECTIVE COATINGS

3040A	3161A	3162A	3163A	1044A	3145A	736B	3077A	1948AB	769C
3170A	3271A	3542A	3243A	1044B	16445A	736C	3097A	1948AC	3049A
2150A	3541A	7852A	8873A	3244A	15525C	3146A	3487A	1948AD	3159A
3430A	3171A	13322B	11403A	4954A	165.4E	31463	3697A	3062A	3519A
11480J	13141K	13332C	11403B	1071A	2055A	3166A	8557A	3125A	3699A
14830A	17041A	133920	11833A	25534A	21185A	8166A	8557B	3258A	8269A
25620A		13392E	20223A		21615A	11656A	8557C	3343A	16409A
		14512E	25613A		25525A	12076A	16497A	11749U	25529A
		14512H				16496A	16527A	14422A	25609A
						16526A		16494A	
						18056B		24198A	
						25466A			
						25596A			

PROTEINS

		124348	160050	5268
		12434C		
		12434E		

IONS

13340C			15705A	1256A	13857B	2029A
			25195A	1256C		13779A
				25196A		13779E

PULSES

		25043A	14934A		15016A		14723B
			15014B		15016B		25088A
			15014D				
			15014E				

PUMPING

		24183A	24554A	14935A	25557A
				14935B	
				14935F	

PUMPS

11170A	81D

PUNCTURE

	24902A

PURIFICATION

25642A	24813A	21134A	24546A	24728A

YRAN 11

			14797B
			16 57A

INFORMATION RETRIEVAL -- DOCUMENT SEARCH NO 8533   DATE   25 NOV 65

ARCH TITLE - CERAMIC REINFORCED COMPOSITES   REQUESTED BY - H EMRICK 36123   PR CUTOFF   5   MAX NO. ALL

ORDER   CONN   WORD NO.

1	AND	3433000	REINFORCEMENT
2	OR	4587500	WHISKERS
3	OR	1467500	FILAMENTS
4	OR	1463000	FIBERS
5	NOT	3250000	POLYMERS
6	AND	1026500	COMPOSITES
7	NOT	1644000	GLASSES
8	NOT	2862000	NITRIDES
9	NOT	791500	CARBIDES
10	NOT	1677500	GRAPHITES
11	NOT	624500	BORIDES
12	NOT	3679050	SINGLE OXIDES
13	NOT	630000	BORON
14	NOT	2572500	MIXED OXIDES

1393A	0126751	013362B	013362C	014034A	014272A	014452A	0146660	015678A	016301C
6783A	017111A	018434A	019250A	020611A	020727A	021312A	022141A	022176A	022238A
22703A	022513A	022532A	022538A	022724A	023569A	023570A	023628A	023949A	024522A
4708A	024734A	024897A	024909A	024909A					

R   35 DOCUMENTS LISTED ABOVE CONTAIN   THE FIRST   14 WORDS.

1442F	0015850	008249A	008672A	011716G	021590A	022271A	024818A		

R   8 DOCUMENTS LISTED ABOVE CONTAIN   THE FIRST   12 WORDS.

1442F	0014420	001442H	0014421	001442J	011689B	012034A	012303A	012500A	014370A
5733A	021398A	022276A	022322A	022533A	024571A				

R   16 DOCUMENTS LISTED ABOVE CONTAIN   THE FIRST   11 WORDS.

1372C

R   1 DOCUMENTS LISTED ABOVE CONTAIN   THE FIRST   10 WORDS.

1917A	010708A	010708B	011877B	011898A	012348A	022020B	022149A	022155A	022511B

R   10 DOCUMENTS LISTED ABOVE CONTAIN   THE FIRST   9 WORDS.

1628A	011462C	012058A	014732A	020551A	022537A	022713A			

R   7 DOCUMENTS LISTED ABOVE CONTAIN   THE FIRST   8 WORDS.

4741A	017367A

R   2 DOCUMENTS LISTED ABOVE CONTAIN   THE FIRST   7 WORDS.

1017JA	000170B	001212A	001442B	002017A	010831A	011689A	011689C	011689E	011870G
2403A	012675B	012675C	012675D	012675E	012675G	012675H	012675L	012675M	013789C
4264A	014386A	014485A	0142323	020725A	024159A				

R   26 DOCUMENTS LISTED ABOVE CONTAIN   THE FIRST   6 WORDS.

000005A	000006A	000007A	000008A	000009A	000010A	000011A	000012A	000013A	000015A
01770C	002203B	000204B	000248C	000264A	000419A	000429A	000475F	000546A	000588B

REPORTS RECEIVED BUT NOT INDEXED

NOT READILY AVAILABLE DURING PROCESSING.
PLEASE REQUEST DOCUMENTS BY ACCESSION NUMBERS.

DOCUMENT CONTROL DATA - R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY *(Corporate author)*	2a. REPORT SECURITY CLASSIFICATION
BOOZ · ALLEN APPLIED RESEARCH, INC. 4733 Bethesda Avenue Bethesda, Maryland 20014	Unclassified
	2b. GROUP

3. REPORT TITLE Mechanization Study of the Aerospace Materials Information Center (Including the Ceramics & Graphites Technical Info Center), Wright-Patterson Air Force Base, Ohio

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)*
Final Report of on-site survey

5. AUTHOR(S) *(Last name, first name, initial)*

G. A. Kershaw, D. Crowder, J. E. Davis, E. G. Loges, E. Merendini, S. M. Thomas

6. REPORT DATE September, 1966	7a. TOTAL NO. OF PAGES 55	7b. NO. OF REFS 2
8a. CONTRACT OR GRANT NO. DSA-7-15489	9a. ORIGINATOR'S REPORT NUMBER(S) 914-1-29	
b. PROJECT NO		
c.	9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)* AD 640 124	
d.		

10. AVAILABILITY LIMITATION NOTICES

Distribution of this Document is unlimited

11. SUPPLEMENTARY NOTES None	12. SPONSORING MILITARY ACTIVITY Defense Supply Agency Defense Documentation Center Cameron Station, Virginia

13. ABSTRACT

By use of the IBM 7094, the Materials Laboratory performs searches and produces the KWIC index section of the "Materials Information Bulletin." Under contract to the Materials Laboratory, thesaurus maintenance, keyword index printing, and data-base updating are performed on the NCR 304 computer at the University of Dayton. The use of deep indexing with links and roles formed the basis of the system; after evaluating the value of links and roles, however, roles were eliminated. A Miracode system just recently purchased will be used to store Materials Laboratory information. Several studies covering specific areas of this system have been conducted. Currently, the Center is sending questionnaires to certain groups after running searches for them to determine the value of each search.

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Digital Computers						
Indexes						
Information Retrieval						
Data						
Analysis						

INSTRUCTIONS

1. ORIGINATING ACTIVITY: Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization (corporate author) issuing the report.

2a. REPORT SECURITY CLASSIFICATION: Enter the overall security classification of the report. Indicate whether "Restricted Data" is included. Marking is to be in accordance with appropriate security regulations.

2b. GROUP: Automatic downgrading is specified in DoD Directive 5200.10 and Armed Forces Industrial Manual. Enter the group number. Also, when applicable, show that optional markings have been used for Group 3 and Group 4 as authorized.

3. REPORT TITLE: Enter the complete report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capitals in parenthesis immediately following the title.

4. DESCRIPTIVE NOTES: If appropriate, enter the type of report, e.g., interim, progress, summary, annual, or final. Give the inclusive dates when a specific reporting period is covered.

5. AUTHOR(S): Enter the name(s) of author(s) as shown on or in the report. Enter last name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement.

6. REPORT DATE: Enter the date of the report as day, month, year, or month, year. If more than one date appears on the report, use date of publication.

7a. TOTAL NUMBER OF PAGES: The total page count should follow normal pagination procedures, i.e., enter the number of pages containing information.

7b. NUMBER OF REFERENCES: Enter the total number of references cited in the report.

8a. CONTRACT OR GRANT NUMBER: If appropriate, enter the applicable number of the contract or grant under which the report was written.

8b, 8c, & 8d. PROJECT NUMBER: Enter the appropriate military department identification, such as project number, subproject number, system numbers, task number, etc.

9a. ORIGINATOR'S REPORT NUMBER(S): Enter the official report number by which the document will be identified and controlled by the originating activity. This number must be unique to this report.

9b. OTHER REPORT NUMBER(S): If the report has been assigned any other report numbers (either by the originator or by the sponsor), also enter this number(s).

10. AVAILABILITY/LIMITATION NOTICES: Enter any limitations on further dissemination of the report, other than those imposed by security classification, using standard statements such as:

(1) "Qualified requesters may obtain copies of this report from DDC."

(2) "Foreign announcement and dissemination of this report by DDC is not authorized."

(3) "U. S. Government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through

_____."

(4) "U. S. military agencies may obtain copies of this report directly from DDC. Other qualified users shall request through

_____."

(5) "All distribution of this report is controlled. Qualified DDC users shall request through

_____."

If the report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known.

11. SUPPLEMENTARY NOTES: Use for additional explanatory notes.

12. SPONSORING MILITARY ACTIVITY: Enter the name of the departmental project office or laboratory sponsoring (paying for) the research and development. Include address.

13. ABSTRACT: Enter an abstract giving a brief and factual summary of the document indicative of the report, even though it may also appear elsewhere in the body of the technical report. If additional space is required, a continuation sheet shall be attached.

It is highly desirable that the abstract of classified reports be unclassified. Each paragraph of the abstract shall end with an indication of the military security classification of the information in the paragraph, represented as (TS), (S), (C), or (U).

There is no limitation on the length of the abstract. However, the suggested length is from 150 to 225 words.

14. KEY WORDS: Key words are technically meaningful terms or short phrases that characterize a report and may be used as index entries for cataloging the report. Key words must be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location, may be used as key words but will be followed by an indication of technical context. The assignment of links, rules, and weights is optional.